Automatic Detection of Major Depressive Disorder via a Bag-of-Behaviour-Words Approach

Kun Qian
Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Japan
qian@p.u-tokyo.ac.jp

Hiroyuki Kuromiya
Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Japan
kuromiya@p.u-tokyo.ac.jp

Zhao Ren
ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany
zhao.ren@informatik.uni-augsburg.de

Maximilian Schmitt
ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany
maximilian.schmitt@informatik.uni-augsburg.de

Zixing Zhang
GLAM–Group on Language, Audio & Music, Imperial College London, UK
zixing.zhang@imperial.ac.uk

Toru Nakamura
Graduate School of Engineering Science, Osaka University, Japan
t-nakamura@sangaku.es.osaka-u.ac.jp

Kazuhiro Yoshiuchi
Department of Stress Sciences and Psychosomatic Medicine, Graduate School of Medicine, The University of Tokyo, Japan
kyoshiuc-tky@umin.ac.jp

Björn W. Schuller*
GLAM–Group on Language, Audio & Music, Imperial College London, UK
schuller@ieee.org

Yoshiharu Yamamoto
Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Japan
yamamoto@p.u-tokyo.ac.jp

ABSTRACT

In recent years, machine learning has been increasingly applied to the area of mental health diagnosis, treatment, support, research, and clinical administration. In particular, using less-invasive wearables combined with the artificial intelligence to monitor, or diagnose the mental diseases has tremendous needs in real practice. To this end, we propose a novel approach for automatic detection of major depressive disorder. Firstly, spontaneous activity physical data are recorded by a watch-type device equipped with an activity monitor. Subsequently, a bag-of-behaviour-words approach is applied to extract higher representations from the raw sensor data in an unsupervised scenario. Finally, a support vector machine is selected as the classifier to make the predictions on screening the major depressive disorder. There are 69 healthy control subjects, and 14 major depressive disorder patients involved in this study. The experimental results demonstrate the effectiveness of the proposed method in a rigorous subject-independent test, which achieves an unweighted average recall at 59.3 % (an accuracy of 66.0 %). This unweighted average recall significantly ($p < .05$, one-tailed z-test) outperforms human hand-crafted features with an unweighted average recall at 53.6 % (an accuracy of 61.7 %).

*Björn W. Schuller is also with the ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany.

ISICDM 2019, August 24–26, 2019, Xi'an, China

CCS CONCEPTS

• **Applied computing → Consumer health**; **Health care information systems**; **Psychology**.

KEYWORDS

Affective Computing, Bag-of-Behaviour-Words, Machine Learning, Major Depressive Disorder, Spontaneous Physical Activity.

ACM Reference Format:
Kun Qian, Hiroyuki Kuromiya, Zhao Ren, Maximilian Schmitt, Zixing Zhang, Toru Nakamura, Kazuhiro Yoshiuchi, Björn W. Schuller, and Yoshiharu Yamamoto. 2019. Automatic Detection of Major Depressive Disorder via a Bag-of-Behaviour-Words Approach. In *The Third International Symposium on Image Computing and Digital Medicine (ISICDM 2019), August 24–26, 2019, Xi'an, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3364836.3364851

1 INTRODUCTION

Major depressive disorder (MDD), defined as a psychiatric disease of having the presence of mood disturbances consistently for more than several weeks [3], affects approximately 3 % of the global population (216 million people) [10]. The previous studies showed that, MDD sufferers could have high risks for cardiovascular morbidity and mortality [11], coronary artery disease, myocardial infarction, and sudden cardiac death [1, 18]. In addition, the risks for suicide will be increased when people are in depression [9, 20]. Therefore, an objective evaluation of depressive mood is vital in terms of the diagnosis and treatment of depressive disorders [14]. Within the fast development of advanced wearable and/or biomedical sensing technologies, it is feasible to collect, and analyse long-term, continuous biomedical signals, which may benefit the area of automatically tracking the mental health status of subjects [16]. In particular, there are tremendous needs in real practice to use the state-of-the-art signal processing and machine learning techniques for analysing the physiological signals, which will lead to the development of portable, less/non-invasive devices for screening, monitoring, and managing a subject's mental health. In this light, we propose a novel bag-of-behaviour-words (BoBW) approach for detection of MDD from spontaneous physical activity (SPA) data [14, 16] recorded via a watch-type device in the subject's daily life.

The main contributions of this work are: Firstly, to the best of our knowledge, it is the first time to introduce a BoBW approach for the analysis of SPA data, by which a subject's behaviour can be mapped into higher representations for further machine learning steps; secondly, there are no human hand-crafted features designed in the proposed paradigm, which can be easily implemented without any specific domain knowledge; thirdly, we investigate the effects of hyper-parameters of the BoBW approach, e. g., the codebook size and the assignment number on the performance of the model. The remainder of this paper is organised as follows: In Section 2, the related work is introduced and discussed. Section 3 describes the database, the methods, and the evaluation metrics used in this study. Experimental results and a discussion are given in Section 4, and Section 5, respectively. Finally, the conclusion and the suggested future research directions are drawn in Section 6.

2 RELATED WORK

In the recent five years, there have been increasingly developing techniques of signal processing and machine learning for the detection of depression [5, 7, 15, 21, 29, 31]. Stratou *et al.* studied the nonverbal behaviour (*affect*, *emotional variability*, and *motor variability*) for the detection of depression and post-traumatic stress disorder [29]. In their work, a Naïve Bayes classifier was selected for making the predictions. Dhall and Goecke introduced a temporally piece-wise fisher vector approach for depression analysis [7]. In the work by Chao *et al.* [5], long short-term memory recurrent neural networks (LSTM RNN) were used to extract sequential information from audio and video features. Pokorny *et al.* introduced the bag-of-audio-words (BoAW) approach into the area of speech emotion recognition [21]. A hybrid system combing the deep convolutional neural network (DCNN) and deep neural network (DNN) models was introduced in [31]. Mustafa *et al.* estimated the heart rate values from facial videos, by which to form a feature vector to classify the depressive patients and the healthy control subjects [15]. Generally, the aforementioned studies had achieved encouraging experimental results for depression analysis and detection. However, there are still some limitations: Firstly, recording long-term audio/video data of the subjects is time-consuming and expensive; secondly, most of the previous studies relied on human hand-crafted features, which may need specific domain knowledge. Motivated by the success of using the SPA data (recorded by the low cost and low energy-consumption wearable sensors) to model the human behaviour in daily life [14, 16, 17], and the successful application of the BoAW approach [23, 25], we introduce, and investigate the efficacy of a novel BoBW approach for depression detection via the SPA data.

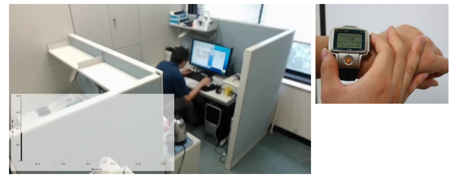

Figure 1: The diagram of the SPA data collection process.

Table 1: The number [#] of instances and participants (shown in brackets) in each data set.

	Train	Dev	Test	Σ
MDD	461 (8)	265 (3)	164 (3)	890 (14)
Healthy	885 (41)	287 (14)	327 (14)	1 499 (69)
Σ	1 346 (49)	552 (17)	491 (17)	2 389 (83)

3 MATERIALS AND METHODS

3.1 Database

This study is approved by the research ethic committees of the University of Tokyo, Japan, and the Teikyo University, Japan. Totally, there are 14 patients with MDD (12 males (M), 2 females (F); age: 34.0±5.7 years, age range: 22-42 years), and 69 healthy control subjects (69 M, 0 F; age: 39.3±9.8 years, age range: 23-58 years) involved. The SPA data of the MDD patients are provided by the Mizonokuchi Hospital of Teikyo University, Kanagawa, Japan. The healthy subjects are full-time office workers from the University of Tokyo, Japan. A watch-type device equipped with an activity monitor (Ruputer ECOLOG, 42 g; Seiko Instruments Inc., Tokyo, Japan), which is analogous in performance to the commercial actigraph (Ambulatory Monitors Inc., Ardsley, NY, USA) [14], was used to record the SPA data (*zero-crossing counts* accumulated for every one minute, see Figure 1 and Figure 2) in the daily life of all participants. Small changes in bodily acceleration (⩾0.01 G/rad/s) can be detected by this sensor. Therefore, behaviours of the participants can be defined by the SPA data in a long term period (e. g., one week, or one month). In the phase of data pre-processing, we cut the whole data recordings of each participant into instances within a same length of 180 min. Specifically, the duration of possible sleep or other non-activity SPA data were excluded. To make this study feasible in real practice, we use a rigid subject-independent evaluation method. The whole database was split into a *train*, a *development* (dev), and a *test* set, respectively. The number participants, and the same as the number of instances in each data set, occupies approximately 60 %, 20 %, and 20 % of the overall database, respectively (see Table 1). Taking the imbalanced characteristic of the SPA data distribution into account, we will use an *upsampling* technique when training the model, by which the instances from the scarce class (MDD in this study) are manually replicated to reach an equal distribution of all classes.

3.2 Statistical Functionals

The recorded SPA data, i. e., the zero-crossing counts, are the raw sensor data. In this study, the SPA data are regarded as the low-level descriptors (LLDs) of the human behaviours, which may reflect the

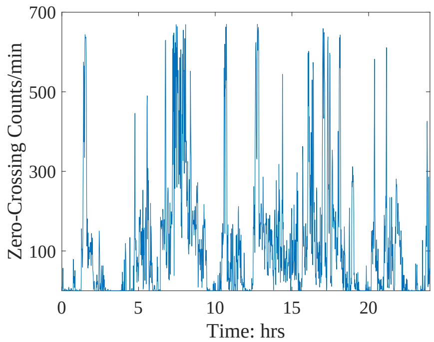

Figure 2: An example of the SPA data recorded by a participant in one day activity.

Table 2: Results (UARs: [%]) achieved by the functionals on the development and the test sets. ↑ indicates the experiments done on the upsampled data.

UAR [%]	Dev	Test
	61.8	53.6
↑	67.8	57.6

Table 3: Results (UARs: [%]) achieved by different parameters of the BoBW approach on the development and the test sets. C_s: Codebook Size; N_a: Assignment Number. ↑ indicates the experiments done on the upsampled data.

	C_s =	10	50	100	500	1 000
N_a = 1	Dev	58.6	61.4	62.9	62.0	61.2
	Test	57.1	54.5	57.9	56.5	58.8
	Dev↑	65.2	67.4	66.5	65.1	65.7
	Test↑	57.7	50.0	57.6	54.8	53.3
N_a = 5	Dev	60.2	61.2	61.5	63.1	60.1
	Test	53.9	47.2	47.2	**59.3**	58.7
	Dev↑	64.5	67.0	67.0	67.0	65.6
	Test↑	52.2	54.2	54.5	57.6	56.8
N_a = 10	Dev	50.0	60.0	61.4	62.1	62.4
	Test	50.0	59.1	45.8	56.2	58.5
	Dev↑	50.0	67.1	67.5	67.6	67.1
	Test↑	50.3	55.7	54.8	56.6	**58.2**

participant's psychological status. It is reasonable to think that, the changes of values of LLDs over a given period of time, may help build the models for learning the participant's behaviour. We use the concept of 'supra-segmental features' [8], to extract the information from a time unit. As a kind of human hand-crafted features, statistical functionals, can map the time series to a scalar value. Motivated by the work in [14, 22], we extract 9 statistical functionals from 180 min LLDs of each instance. These functionals are the values of *maximum*, *minimum*, *mean*, *range*, *standard deviation*, *skewness*, *kurtosis*, *slope*, and *bias* of the linear regression approximation for the LLDs over a time unit (180 min).

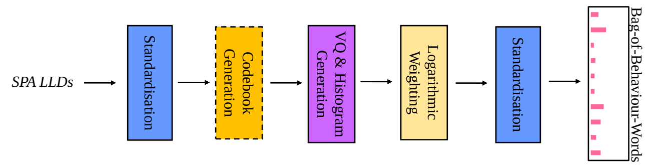

Figure 3: The general diagram of the proposed BoBW approach.

3.3 Bag-of-Behaviour-Words Approach

The *bag-of-words* (BoW) approach is known from *natural language processing* [30], which can be referred to the early description in [12]. Particularly, in the application of *speech emotion recognition* [21, 26], and the area of health care [13, 25], the BoAW approach achieved numerous excellent results. Motivated by the success of the BoAW method in aforementioned studies, we propose the *bad-of-behaviour-words* (BoBW) approach. The general diagram of the BoBW approach is shown in Figure 3. In the BoBW approach, the SPA data (in 180 min) will be firstly divided into *frames* within a length of 10 min, and an overlap of 5 min of each neighbour. These frame-based LLDs will be put into a phase of *vector quantisation* (VQ), which is done by using a *codebook* of template LLDs ('behaviour words') learnt previously from the training data. Usually, *K-means clustering* is employed for the codebook generation [19]. However, similar results can be reached by *random sampling* [24] following the initialisation step of *K-means++ clustering* [2], in which the far-off LLDs are prioritised for the iterative selection of the cluster centroids. The N_a words with the lowest *Euclidean distance* are considered instead of assigning each LLD only to the most similar word in the codebook, which was demonstrated to improve the robustness of the method [26]. The *term-frequency histogram* is generated by counting the *term-frequencies*, i. e., the number of each behaviour word that has been chosen as the nearest neighbour for the LLDs in one SPA instance. Finally, the logarithm (with a bias of one) is taken from the word frequencies in order to compress the range of values in the resulting histogram.

3.4 Support Vector Machines

In this study, we use the popular and standard machine learning model, i. e., support vector machines (SVM) [6] as the classifier. For an SVM classifier, the process of training is to find the best hyperplane that can maximise the separation between classes. When making a decision, the instances will be mapped onto a multidimensional space firstly, and the predictions are given based on which side of the gap these instances fall onto.

3.5 Evaluation Metrics

As a general evaluation metric in this study, the unweighted average recall (UAR) [28], i. e., the averaged recall of each class, is used due to the imbalanced characteristic of the SPA data distribution (see Table 1). Besides, some other metrics, including *accuracy*, *sensitivity*, *specificity*, *precision*, *recall*, *F1-measure*, and *the area under the curve* (AUC) of the receiver operating characteristic will be provided. To give the significance level test when comparing results, a one-tailed z-test is used.

Table 4: Other evaluation metrics ([%]) achieved by the best models based on the functionals and the BoBW approaches on the test set. ↑ indicates the experiments done on the upsampled data.

	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-measure	AUC
func.	61.7	29.3	78.0	40.0	29.3	33.8	44.8
BoBW	66.0	39.0	79.5	48.9	39.0	43.4	37.8
p-value	-	$p < .05$	-	$p < .005$	$p < .05$	$p < .001$	$p < .05$
func.↑	59.1	53.0	62.1	41.2	53.0	46.4	59.9
BoBW↑	63.1	43.3	73.1	44.7	43.3	44.0	64.2
p-value	-	$p < .05$	$p < .002$	-	$p < .05$	-	-

Table 5: Confusion matrices (normalised: [%]) of the best models based on the functionals and the BoBW approaches on the test set. ↑ indicates the experiments done on the upsampled data.

(a) func.

Pred ->	MDD	Healthy
MDD	29.3	70.7
Healthy	22.0	78.0

(b) BoBW

Pred ->	MDD	Healthy
MDD	39.0	61.0
Healthy	20.5	79.5

(c) func.↑

Pred ->	MDD	Healthy
MDD	53.0	47.0
Healthy	37.9	62.1

(d) BoBW↑

Pred ->	MDD	Healthy
MDD	43.3	56.7
Healthy	26.9	73.1

4 EXPERIMENTAL RESULTS

4.1 Experimental Setup

The statistical functionals were extracted in an environment of Matlab R2018b by MathWorks (Natick, MA, USA). For the BoBW approach, we use the openXBOW toolkit [27]. The popular open source toolkit LIBSVM [4] was used to implement the SVM classifier. The kernels of the SVM classifier were selected from *linear*, *polynomial*, *sigmoid*, and *radial basis function (RBF)*. The C-value for the SVM classifier was optimised by a grid-search strategy from $\{10^{-5}, 10^{-4}, \ldots, 10^{4}, 10^{5}\}$. All the hyper-parameters of the classifiers were optimised on the development set, and applied to the test set. For the random sampling of the LDDs in BoBW, we used the default random seed in the openXBOW toolkit. In this study, we also investigated the effects of hyper-parameters of the BoBW approach by setting the *codebook size* (C_s), and the *assignment number* (N_a) from a grid of {10, 50, 100, 500, 1 000}, and {1, 5, 10}, respectively. Both of the features by the method of functionals and BoBW, are standardised to eliminate outliers before fed into the SVM classifier.

4.2 Results

The experimental results (UARs, in [%]) achieved by the functionals and the BoBW approaches are shown in Table 2, and Table 3, respectively. The results for the development set are the ones of the optimised models. We can find that, by upsampling, the final performance of the model in the development set can be improved in most of the cases. Without upsampling, BoBW (within an UAR of 59.3 %) can outperform functionals (within an UAR of 53.6 %) significantly ($p < .05$, one-tailed z-test). On one hand, upsampling helps enhance the performance of the funtionals on both of the development and the test sets. On the other hand, the best performance of the BoBW method has a slight decrease (from 59.3 % to 58.2 % of UAR) on the test set. Table 4 illustrates some other metrics for evaluation of the models. Generally, the proposed BoBW approach outpefroms the functionals when looking at the accuracy. The BoBW-based model reaches 66.0 % of the accuracy without upsampling while the functionals-based model gets 61.7 %. For most of the other metrics (except the specificity and the AUC), BoBW shows its better performance than functionals when using the original data. Nevertheless, this superior characteristic of BoBW to functionals will not be lasting when using the upsampled data. The confusion matrices of the best results achieved by the functionals and the BoBW approaches are given in Table 5. It can be found that, for both of the two methods, i. e., functionals and BoBW, upsampling can improve the recall of MDD. However, the corresponding recall of the healthy control will be decreased by adding more MDD instances.

5 DISCUSSION

Feature engineering is an essential step in conventional machine learning framework. Nevertheless, designing efficient and robust features for specific tasks is usually expensive, tedious, and time-consuming. It is encouraging to find that, the proposed BoBW approach performs well in detection of MDD via the SPA data even in an extremely imbalanced data distribution scenario. In addition, compared with the conventional human hand-crafted features, i. e., the functionals, the BoBW has a better performance in this study. Particularly, one consistent finding in the results of the current work is that, the traditional functionals are more vulnerable to the imbalanced data distribution than the BoBW when using the data without upsampling. However, we should note that, the hyper-parameters of the BoBW approach, i. e., the codebook size and the assignment number, do really have effects on the final performance (see Table 3). Currently, these hyper-parameters are tuned empirically, which depends on specific tasks, or applications. To this end, one direction of future study could be exploring some automatic methods on setting suitable hyper-parameters for the BoBW approach. When evaluating other metrics, we may find that, for both of the functionals and the BoBW, accuracy and specificity are promising (see Table 4). However, sensitivity, precision, recall, and F1-measure, are still lacking on an improvement. One reasonable explanation could be that, current database contains very limited instances from MDD patients, which may lead the trained models perform better in excluding the healthy control rather than detecting the MDD patients. This could also be a reason why the recall of MDD can be improved when adding some more MDD instances by upsampling (see Table 5). The AUC of the model trained by the functionals is better than the one trained by the BoBW without data upsampling (see Table 4). After data upsampling, the BoBW outperforms the functionals with a best AUC (0.642, see Table 4) achieved in this study. Therefore, in future work, we may consider involving more advanced data augmentation methods to improve the baseline.

6 CONCLUSION

In this study, we proposed a novel BoBW approach for detection of MDD from the SPA data recorded in the subject's daily life. There were no human hand-crafted features involved in the paradigm, i. e., higher representations were extracted from the raw SPA data via the BoBW approach in an unsupervised scenario. Experimental results demonstrated the efficacy of the proposed method, which might benefit the development of a smart device in terms of using the SPA data for screening and monitoring the MDD patients. In future work, we will investigate more advanced features extracted from the deep learning models combined with the BoBW approach.

ACKNOWLEDGMENTS

This work was partially supported by the Grants-in-Aid for Scientific Research (17H00878) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and the EU's Horizon H2020 Marie Skłodowska-Curie grant agreement No. 766287 (TAPAS). The authors would like to thank all the colleagues involved in the collection of the SPA data. The first author would like to thank Ms. Xiao (Lexie) Li for her love and support.

REFERENCES

[1] Gail A Alvares, Daniel S Quintana, Ian B Hickie, and Adam J Guastella. 2016. Autonomic nervous system dysfunction in psychiatric disorders and the impact of psychotropic medications: A systematic review and meta-analysis. *Journal of Psychiatry & Neuroscience* 41, 2 (2016), 89–104.

[2] D. Arthur and S. Vassilvitskii. 2007. K-means++: The advantages of careful seeding. In *Proc. ACM-SIAM SODA*. ACM, New Orleans, LA, USA, 1027–1035.

[3] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.* American Psychiatric Association, Arlington, VA, USA.

[4] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2 (2011), 27:1–27:27. Issue 3. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[5] Linlin Chao, Jianhua Tao, Minghao Yang, and Ya Li. 2015. Multi task sequence learning for depression scale prediction from video. In *Proc. ACII*. IEEE, Xi'an, P. R. China, 526–531.

[6] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning* 20, 3 (1995), 273–297.

[7] Abhinav Dhall and Roland Goecke. 2015. A temporally piece-wise fisher vector approach for depression analysis. In *Proc. ACII*. IEEE, Xi'an, P. R. China, 255–259.

[8] Florian Eyben. 2015. *Real-time Speech and Music Classification by Large Audio Feature Space Extraction.* Springer International Publishing, Cham, Switzerland. Doctoral Thesis.

[9] Xinyu Fang, Chen Zhang, Zhiguo Wu, Daihui Peng, Weiping Xia, Jingjing Xu, Chenglei Wang, Lvchun Cui, Jia Huang, and Yiru Fang. 2018. Prevalence, risk factors and clinical characteristics of suicidal ideation in Chinese patients with depression. *Journal of Affective Disorders* 235 (2018), 135–141.

[10] GBD 2015 Mortality and Causes of Death Collaborators. 2016. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* 388, 10053 (2016), 1545–1602.

[11] David L Hare, Samia R Toukhsati, Peter Johansson, and Tiny Jaarsma. 2013. Depression and cardiovascular disease: A clinical review. *European Heart Journal* 35, 21 (2013), 1365–1372.

[12] Zellig S Harris. 1954. Distributional structure. *Word* 10, 2-3 (1954), 146–162.

[13] Jyoti Joshi, Roland Goecke, Sharifa Alghowinem, Abhinav Dhall, Michael Wagner, Julien Epps, Gordon Parker, and Michael Breakspear. 2013. Multimodal assistive technologies for depression diagnosis and monitoring. *Journal on Multimodal User Interfaces* 7, 3 (2013), 217–228.

[14] Jinhyuk Kim, Toru Nakamura, Hiroe Kikuchi, Kazuhiro Yoshiuchi, Tsukasa Sasaki, and Yoshiharu Yamamoto. 2015. Covariation of depressive mood and spontaneous physical activity in major depressive disorder: Toward continuous monitoring of depressive mood. *IEEE Journal of Biomedical and Health Informatics* 19, 4 (2015), 1347–1355.

[15] Aamir Mustafa, Shalini Bhatia, Munawar Hayat, and Roland Goecke. 2017. Heart rate estimation from facial videos for depression analysis. In *Proc. ACII*. IEEE, San Antonio, TX, USA, 498–503.

[16] Toru Nakamura, Ken Kiyono, Herwig Wendt, Patrice Abry, and Yoshiharu Yamamoto. 2016. Multiscale analysis of intensive longitudinal biomedical signals and its clinical applications. *Proc. IEEE* 104, 2 (2016), 242–261.

[17] Toru Nakamura, Ken Kiyono, Kazuhiro Yoshiuchi, Rika Nakahara, Zbigniew R Struzik, and Yoshiharu Yamamoto. 2007. Universal scaling law in human behavioral organization. *Physical Review Letters* 99, 13 (2007), 138103.

[18] Amanda Nicholson, Hannah Kuper, and Harry Hemingway. 2006. Depression as an aetiologic and prognostic factor in coronary heart disease: A meta-analysis of 6362 events among 146 538 participants in 54 observational studies. *European Heart Journal* 27, 23 (2006), 2763–2774.

[19] S. Pancoast and M. Akbacak. 2012. Bag-of-Audio-Words Approach for Multimedia Event Classification. In *Proc. INTERSPEECH*. ISCA, Portland, OR, USA, 2105–2108.

[20] Seon-Cheol Park, Min-Soo Lee, Sang Woo Hahn, Tian-Mei Si, Shigenobu Kanba, Mian-Yoon Chong, Chee Kok Yoon, Pichet Udomratn, Adarsh Tripathi, Norman Sartorius, Naotaka Shinfuku, Margarita M. Maramis, and Yong Chon Park. 2016. Suicidal thoughts/acts and clinical correlates in patients with depressive disorders in Asians: Results from the REAP-AD study. *Acta Neuropsychiatrica* 28, 6 (2016), 337–345.

[21] Florian B Pokorny, Franz Graf, Franz Pernkopf, and Björn W Schuller. 2015. Detection of negative emotions in speech signals using bags-of-audio-words. In *Proc. ACII*. IEEE, Xi'an, P. R. China, 879–884.

[22] Kun Qian, Christoph Janott, Vedhas Pandit, Zixing Zhang, Clemens Heiser, Winfried Hohenhorst, Michael Herzog, Werner Hemmert, and Björn Schuller. 2017. Classification of the Excitation Location of Snore Sounds in the Upper Airway by Acoustic Multi-Feature Analysis. *IEEE Transactions on Biomedical Engineering* 64, 8 (2017), 1731–1741.

[23] Kun Qian, Maximilian Schmitt, Christoph Janott, Zixing Zhang, Clemens Heiser, Winfried Hohenhorst, Michael Herzog, Werner Hemmert, and Björn Schuller. 2019. A Bag of Wavelet Features for Snore Sound Classification. *Annals of Biomedical Engineering* 47, 4 (2019), 1000–1011.

[24] S. Rawat, P. F. Schulam, S. Burger, D. Ding, Y. Wang, and F. Metze. 2013. Robust audio-codebooks for large-scale event detection in consumer videos. In *Proc. INTERSPEECH*. ISCA, Lyon, France, 2929–2933.

[25] Maximilian Schmitt, Christoph Janott, Vedhas Pandit, Kun Qian, Clemens Heiser, Werner Hemmert, and Björn Schuller. 2016. A Bag-of-Audio-Words Approach for Snore Sounds' Excitation Localisation. In *Proc. ITG Speech Communication*. VDE, Paderborn, Germany, 230–234.

[26] M. Schmitt, F. Ringeval, and B. Schuller. 2016. At the Border of Acoustics and Linguistics: Bag-of-Audio-Words for the Recognition of Emotions in Speech. In *Proc. INTERSPEECH*. ISCA, San Francisco, CA, USA, 495–499.

[27] M. Schmitt and B. W. Schuller. 2017. openXBOW-Introducing the Passau Open-Source Crossmodal Bag-of-Words Toolkit. *Journal of Machine Learning Research* 18, 96 (2017), 1–5.

[28] Björn Schuller, Stefan Steidl, and Anton Batliner. 2009. The INTERSPEECH 2009 emotion challenge. In *Proc. INTERSPEECH*. ISCA, Brighton, UK, 312–315.

[29] Giota Stratou, Stefan Scherer, Jonathan Gratch, and Louis-Philippe Morency. 2013. Automatic nonverbal behavior indicators of depression and PTSD: Exploring gender differences. In *Proc. ACII*. IEEE, Geneva, Switzerland, 147–152.

[30] Felix Weninger, Pascal Staudt, and Björn Schuller. 2013. Words that fascinate the listener: Predicting affective ratings of on-line lectures. *International Journal of Distance Education Technologies* 11, 2 (2013), 110–123.

[31] Le Yang, Dongmei Jiang, Wenjing Han, and Hichem Sahli. 2017. DCNN and DNN based multi-modal depression recognition. In *Proc. ACII*. IEEE, San Antonio, TX, USA, 484–489.